IITA Institute For Information Technology Applications

US Air Force Academy

Advanced Usability Evaluation Methods

Terence S. Andre, Lt Col, USAF

Margaret Schurig, Human Factors Design Specialist,

The Boeing Co.

Institute for Information Technology Applications

United States Air Force Academy, Colorado

IITA Technical Report

TR-07-2

April 2007

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **APR 2007**	2. REPORT TYPE	3. DATES COVERED **00-00-2007 to 00-00-2007**
4. TITLE AND SUBTITLE **Advanced Usability Evaluation Methods**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Institute for Information Technology Applications,HQ USAFA/DFPS,2354 Fairchild Drive Suite 6L16D,USAF Academy,CO,80840-6258**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR’S ACRONYM(S)
		11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		

14. ABSTRACT

The Behavioral Sciences and Leadership Department at the United States Air Force Academy (USAFA) developed a low-cost human-computer interaction (HCI) laboratory in 2004. Since that time, the lab has grown into a teaching laboratory introducing HCI concepts to cadets in the Behavioral Sciences-Human Factors option as well as cadets in the Systems Engineering-Human Systems concentration. The HCI lab exposed cadets to contemporary methods and tools used in usability evaluation. The purpose of this final report is to document two studies recently conducted in the HCI laboratory. The first study examined the use of eye tracking as an advanced technique in determining the attentional focus of an evaluator watching a recorded usability highlight video. Current usability evaluation recording technology provides the usability practitioner with the capability to record audio, video of the user, and desktop screen activity in a ?picture-in-picture? (PIP) format, allowing the evaluator to observe the interface screen and the human user simultaneously. The research question in the first study focused on how best to present the PIP video that is often displayed along with the desktop screen capture. A total of 16 undergraduate evaluators were used, with 8 having no experience and 8 having 20 hours of experience from an HCI course. In addition, 6 usability practitioners were used to compare to undergraduate experiences. Results showed that opacity levels of the PIP video did not influence the number of usability problems found for all three groups. All evaluators did focus more on the higher opacity PIP video, but this did not appear to influence their evaluation. In the second study, instructors in the Behavioral Sciences and Leadership Department were interested in examining the changes in a student?s technique of identifying usability problems while using the HCI laboratory. Practitioners in the usability field have noted that experience contributes to the quality of usability problem reports, especially when that experience includes exposure to a framework for doing usability evaluation. Thirteen students in an undergraduate HCI course participated in this study during the Fall 2006 semester. During a pre-and post-assessment, we collected several measures in order to quantify any changes experienced by the students as they logged usability problems. These measures included attention focus, number of problems identified, word count, and use of HCI terms in describing usability problems. Results showed that the metrics of number of usability problems identified and the use of HCI technical terms were particularly sensitive to changes over the semester. The studies discussed in this report provide both researchers and practitioners a way of quantifying the attentional focus of an evaluator using a particular method. In addition results provide those who teach HCI methods a way to measure growth in experience throughout a course.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**18**	

The views expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of the Institute for Information Technology Application, the Department of the Air Force, the Department of Defense or the U.S. Government.

Comments pertaining to this report are invited and should be directed to:

Sharon Richardson
Director of Conferences and Publication
Institute for Information Technology Applications
HQ USAFA/DFPS
2354 Fairchild Drive, Suite 6L16D
USAF Academy CO 80840-6258
Tel. (719) 333-2746; Fax (719) 333-2945
Email: sharon.richardson@usafa.af.mil

Table of Contents

List of Figures

Advanced Usability Evaluation Methods

Abstract

The Behavioral Sciences and Leadership Department at the United States Air Force Academy (USAFA) developed a low-cost human-computer interaction (HCI) laboratory in 2004. Since that time, the lab has grown into a teaching laboratory introducing HCI concepts to cadets in the Behavioral Sciences-Human Factors option as well as cadets in the Systems Engineering-Human Systems concentration. The HCI lab exposed cadets to contemporary methods and tools used in usability evaluation. The purpose of this final report is to document two studies recently conducted in the HCI laboratory. The first study examined the use of eye tracking as an advanced technique in determining the attentional focus of an evaluator watching a recorded usability highlight video. Current usability evaluation recording technology provides the usability practitioner with the capability to record audio, video of the user, and desktop screen activity in a “picture-in-picture” (PIP) format, allowing the evaluator to observe the interface screen and the human user simultaneously. The research question in the first study focused on how best to present the PIP video that is often displayed along with the desktop screen capture. A total of 16 undergraduate evaluators were used, with 8 having no experience and 8 having 20 hours of experience from an HCI course. In addition, 6 usability practitioners were used to compare to undergraduate experiences. Results showed that opacity levels of the PIP video did not influence the number of usability problems found for all three groups. All evaluators did focus more on the higher opacity PIP video, but this did not appear to influence their evaluation.

In the second study, instructors in the Behavioral Sciences and Leadership Department were interested in examining the changes in a student’s technique of identifying usability problems while using the HCI laboratory. Practitioners in the usability field have noted that experience contributes to the quality of usability problem reports, especially when that experience includes exposure to a framework for doing usability evaluation. Thirteen students in an undergraduate HCI course participated in this study during the Fall 2006 semester. During a pre-and post-assessment, we collected several measures in order to quantify any changes experienced by the students as they logged usability problems. These measures included attention focus, number of problems identified, word count, and use of HCI terms in describing usability problems. Results showed that the metrics of number of usability problems identified and the use of HCI technical terms were particularly sensitive to changes over the semester.

The studies discussed in this report provide both researchers and practitioners a way of quantifying the attentional focus of an evaluator using a particular method. In addition, results provide those who teach HCI methods a way to measure growth in experience throughout a course.

Study 1 – Eye Tracking of Evaluators

Introduction

Usability testing has become a common practice in industry due to the importance of making usable software interfaces and the availability of methods and tools (Andre, Hartson, Belz, & McCreary, 2001). Almost every company that develops a desktop or web interface product uses some level of usability testing and evaluation to improve their product before it is launched. Better user interfaces can often become the distinguishing feature that provides a competitive advantage. A common technique in usability testing is the use of screen capture and audio/video recordings of a subject as a way to identify usability problems and errors. Just a few years ago, digital recording of usability testing sessions were reserved for companies like Microsoft, Oracle, Sun Microsystems, and IBM. These companies have high-end expensive laboratories with a dedicated staff who conduct the usability testing on all their products. Digital recording is now readily available through desktop software to almost anyone who wants to create a usability evaluation laboratory. All that is needed is a computer, software, and a web camera.

The recent advances in usability recording technology allow practitioners to create multimedia recordings from a usability evaluation session where audio, video of the user, and the desktop screen activity are integrated into a picture-in-picture (PIP) video. The video shows the desktop activity in the largest area of the screen with a small PIP video of the user in the lower right corner. The purpose of the small PIP video of the user is to capture nonverbal cues that can sometimes lead to discovering a usability problem that is not obvious from just focusing on the desktop actions. Patterson (1983) suggests that nonverbal cues can be representative of the true feelings and attitudes of a person as they accomplish a task. Research in the area of video conferencing has indicated that nonverbal cues can enhance verbal communication which is used through a participant's introspection of their performance on a designated task (Argyle, 1972; Argyle & Dean, 1965; Argyle, Lalljee, & Cook, 1968; Kendon, 1967). In the field of usability evaluation, the specific benefits of including PIP video of user nonverbal cues is undetermined. Usability practitioners typically recognize that effective usability evaluation analysis must always include the desktop screen activity and the user audio. In remote usability testing, practitioners are usually limited to desktop screen activity and audio anyway and this becomes the default standard. Still, questions remain as to the benefit of PIP video and how best to integrate it when it is available. For example, what size of the screen video should the PIP occupy? Is it important that the PIP is somewhat translucent so that the observer can see through the PIP video to what is happening on the desktop? These are just some of the questions that have not been answered for usability practitioners who now have this capability readily available to them. Because it is easy to integrate this technology, it does not necessarily mean that usability of a product is improved because a practitioner uses all of it.

In order to answer some of our research questions, we needed to quantify what evaluators are specifically looking at when watching usability recordings. A tool that has become more readily available and relatively easy to use is eye tracking equipment. Eye tracking has traditionally been used in human performance studies in aviation and to determine potential usability problems in a software interface. Eye movements usually indicate a person's spatial focus of attention on a display (Goldberg and Kotval, 1999).

Goldberg and Wichansky (2003) and Lin and Zhang (2003) have documented the potential eye tracking measures that appear to be more sensitive to different interface designs. Eye tracking has not been used to determine what the evaluator is looking at when viewing usability recordings.

We conducted an earlier study in our laboratory that focused on the presence/absence of the PIP video in usability lab recordings (Long, Styles, Andre, and Malcom, 2005). One group of practitioners watched several pre-recorded usability videos with the PIP video present while a second group of practitioners watched the same videos without the PIP video present. The common elements available to both groups were the screen capture of the desktop activity and the audio (verbal protocol) of the user. Thus, the unique aspect was whether the PIP video of the user nonverbal activity was present or absent. Results showed that the presence/absence of the PIP video did not affect the average number of problems found between the two groups. Evaluators without the PIP video had a much larger range of problems identified (i.e., higher variation) than the evaluators who had the PIP video available to them. Long, Styles, Andre, and Malcom (2005) concluded that there were potentially more attentional demands on the PIP video group (i.e., audio, desktop screen capture, and PIP video) that limited the possible range of problems identified. The PIP video appeared to help confirm usability problems, but did not lead to finding significantly more or less usability problems.

This study focused on the quality of the PIP video in terms of opacity. Opacity is the quality of an object that makes it impervious to rays of light passing through it. In usability evaluation recordings, opacity is often set to something less than 100 percent in order to let the evaluator see both the user activity (PIP video) and the desktop screen capture behind the PIP video. Our research objectives were two-fold. First, how much attention is given to the PIP video of user actions? Second, what impact does the opacity of the PIP video have on identifying usability problems? In addition, we also examined these same questions across evaluator experience.

Method

Participants

We used 16 undergraduate students from the United States Air Force Academy and 6 usability practitioners from industry in this study. Eight of the undergraduate students had no previous experience in usability evaluation (novice undergrads) while the other eight were enrolled in a human-computer interaction (HCI) course (experienced undergrads). The HCI course gave students approximately 20 hours of experience in usability evaluation. The six usability practitioners (experienced practitioners) had an average of 10 years experience in HCI.

Apparatus

The Eye-gaze Response Interface Computer Aid (ERICA) eye tracker and Gazetracker analysis software were used to determine the evaluator’s focus of attention when viewing usability videos.

Procedure

All 22 evaluators completed a web-based training program on identifying usability problems in a software interface. The novice undergrads and experienced practitioners watched sample videos at the end of their training showing an expert evaluator identifying problems. Experienced undergrads did not have to watch these sample videos since they had recently seen these in their HCI course. Each group of evaluators watched two scripted usability recordings. These scripted usability recordings were created ahead of time with an actor who used the same interface but encountered slightly different problems in each recording. The usability problems were different enough in each recording so that evaluators would not recognize the exact same problem. The usability recordings differed in the level of opacity of the PIP video. One recording was set with a PIP video opacity of 50 percent while the other recording was set to 100 percent as shown in Figure 1.

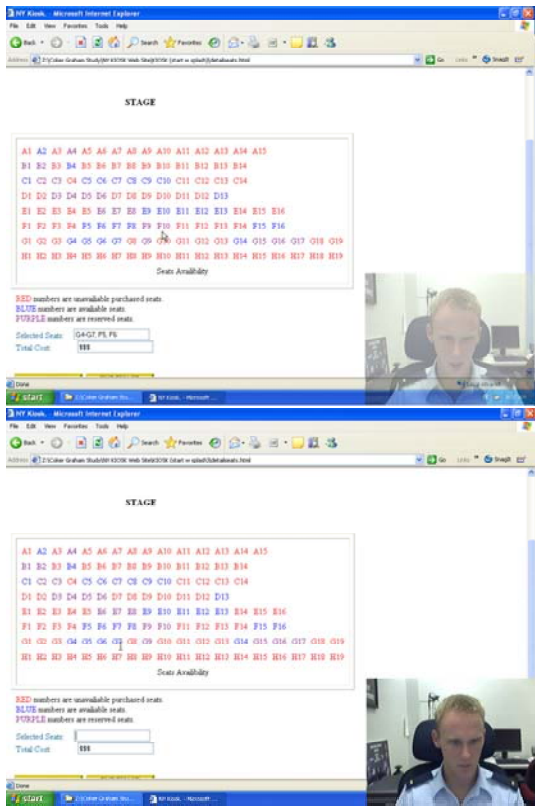

Figure 1: PIP video at 50 percent opacity (top) and 100 percent opacity (bottom)

Presentation of the two usability recording videos were counterbalanced in order to reduce learning effects. Evaluators indicated a usability problem from the video by pressing the space bar, which was then captured along with the eye tracking data.

Results

As expected, all evaluators spent more time focused on the desktop activity than the PIP video as shown in Figure 2 [$F(1,19)=487.10$, $p<.0001$]. Novice undergrads on average

focused on the desktop activity 85.78% of the time compared to 8.00% for the PIP video. Experienced undergrads focused on the desktop activity 76.47% of the time compared to 15.15% for the PIP video. Finally, experienced practitioners focused on the desktop activity 89.77% of the time compared to 3.74% for the PIP video. Note that the percentages do not add up to 100% due to evaluators looking off the screen (approximately 7-9% of the time).

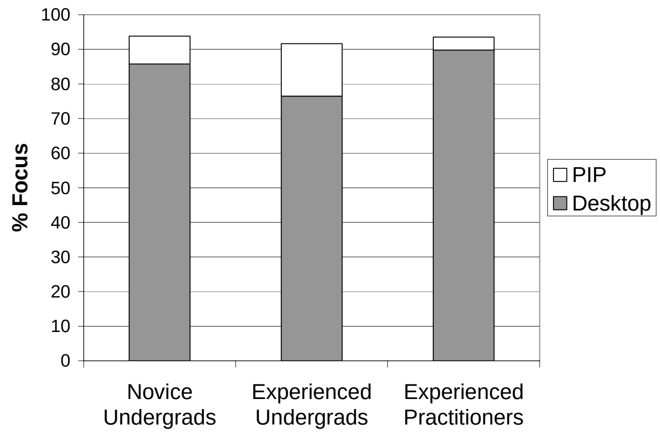

Figure 2: Percent focus on desktop activity vs. PIP video with respect to level of experience

Results also showed an interaction for the percent time focused on PIP video across experience levels [$F(2,19)=4.54$, $p=.024$]. Experienced undergrads spent over four times that of experienced practitioners looking at the PIP video (Bonferroni, $p=.015$).

Figure 3 shows that all evaluators spent more time focusing on the PIP video when it was set to 100% opacity [$F(1,19)=12.51$, $p=.002$].

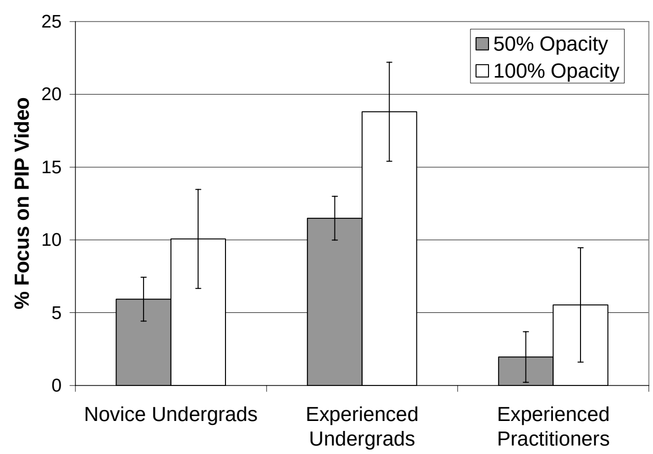

Figure 3: Percent focus on PIP video with respect to opacity setting (50% vs. 100%)

Figure 4 shows that the varying levels of PIP video opacity did not significantly effect the number of usability problems identified [$F(1,19)=0.898$, $p>.10$]. Within each group, evaluators found approximately the same number of problems when watching videos with 50% and 100% opacity. However, there was an effect for experience level [$F(2,19)=5.03$, $p=.018$]. Experienced undergrads found significantly more usability problems than novice undergrads (23.5 vs. 15.1) (Bonferroni, $p=.019$). Experienced undergrads also found more usability problems than experienced practitioners (23.5 vs. 17.2), but this difference was not significant (Bonferroni, $p>.10$).

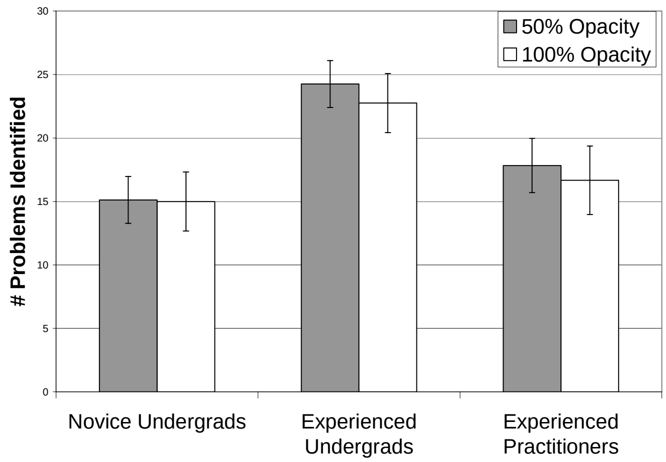

Figure 4: Mean number of usability problems identified with respect to opacity setting (50% vs. 100%)

Discussion

The results clearly show that opacity has no significant effect on an evaluator's ability to perform a usability analysis. Experienced undergraduate evaluators shift their attention more to the PIP video and find more usability problems than novice undergrads (significant). Numerically, this same trend was observed between experienced undergrads and experienced practitioners, but this trend was not significant probably because of the small number of experienced practitioners. Experienced undergrads were probably more aware of the PIP video and interested in finding all kinds of usability problems because of their recent course training. In their course training, experienced undergrads learned about the importance of user "nonverbal" actions. It is possible that many of the experienced undergraduates were eager to implement recently taught techniques of identifying usability problems and therefore paid attention to the PIP video, and reported more problems.

Based on these results and the results from the Long et al. (2005) study, we can support the conclusion that desktop screen activity and audio of the user are essential elements for evaluators finding usability problems. PIP video of nonverbal behaviour appears to have an effect on the evaluator's attentional focus, but without any significant impact on the number of problems they identify. The benefit of PIP video in terms of usability evaluation is still unknown, at least in terms of the context of this study. When a user is particularly expressive in verbal protocol during a session, it is quite possible that PIP

video is unnecessary for evaluators. The PIP video, in some cases, may serve to confirm a usability problem that is observed primarily through desktop activity and user verbal protocol (audio). When the user is not particularly expressive in verbal protocol, the PIP video may become more valuable.

A secondary finding of this study is that there are real differences in outcomes when considering the experience levels of evaluators as measured by attention focus and number of problems reported. Although the experience differences are not surprising, the fact that eye tracking analysis provides a quantifiable difference is particularly beneficial to the field. These experience differences also led us to conduct a second study where we examined the specific changes that occur in undergraduate students as they learn how to do usability evaluation.

Study 2 – Measuring Changes in Usability Experience

Introduction

Our undergraduate HCI laboratory at the United States Air Force Academy is coupled with a specific course that provides a teaching laboratory of usability concepts and methods for cadets. With the help of the TechSmith Morae™ software, we were able to build a fully digital teaching laboratory for cadets. We developed a brand new HCI course using the Interaction Design textbook by Preece, Rodgers, and Sharp (2002). The course presents basic components of HCI concepts, theory, and practice from a user-centered perspective. A central theme is on design and evaluation as highly iterative and connected processes using a usability engineering life cycle framework.

After the first offering of the course in the Fall 2005 semester, we started thinking about ways to measure how effective our teaching laboratory was at developing usability expertise in our students. We had traditional assessments such as quizzes, exams, and project reports, but felt we needed to measure some output of actually “doing” usability evaluation. Specifically, we were interested in changes in their technique of identifying usability problems while using the HCI laboratory.

Hartson, Andre, and Williges (2003) have noted that usability problem reporting is often adhoc and based upon whatever the evaluator thinks of at that time. Standards for using usability evaluation methods and definitions of measures have considerable variation in the HCI field (Gray and Salzman, 1998; Sears, 1997). There are also experience differences between novice and expert evaluators when describing usability problems (Andre, Graham, Coker, & Schurig, 2006). Experienced usability practitioners typically find more usability problems than novice evaluators because they know what they are looking for and are familiar with common design standards. In addition, usability practitioners refine their technique over time as they experience a variety of different usability issues across different applications.

Based on our background research, we focused on measures that could potentially quantify the experience changes that occurred in students taking an undergraduate HCI course. Our process yielded the following measures that were readily available using existing lab resources:

- Attention focus collected by eye tracking equipment (e.g., how much do evaluators look at the desktop activity vs. the picture-in-picture video of the user)
- Number of usability problems identified (a traditional usability measure)
- Average word count in describing a usability problem
- Use of a specific set of HCI terms in describing usability problems

Our research was exploratory in nature in order to develop objective measures that could be used in future courses. We did expect to see differences in some of these measures over time as we tested students early in the semester and then at the end of the semester. Specifically, we expected that at the end of the semester students would find more usability problems, use fewer words to describe each problem, and use a greater percentage of HCI terms in their descriptions. Our previous work had shown us that expert practitioners look at the PIP video of the user slightly less than novice evaluators (Andre et al., 2006). We did not expect a significant difference in this attention focus measure but did want to note any trends.

Method

Participants

We used 13 undergraduate students from the United States Air Force Academy. These 13 students were enrolled in the HCI course during the Fall 2006 semester. These students also had courses in human factors, cognitive psychology, research methods, and engineering psychology. None of the students had been exposed to formal HCI principles in a complete course.

Apparatus

Students used a paper-based critical incident report form to log the usability problems observed during both the pre- and post-assessments. In addition, we recorded eye tracking of each student using the Eye-gaze Response Interface Computer Aid (ERICA) to determine the student's focus of attention.

Procedure

The 13 students completed a web-based training program on identifying usability problems in a software interface at the beginning of the semester and were assessed by watching a pre-recorded session of a user interacting with a fictitious web application for buying online theatre tickets. Students could see a PIP video of the user, synchronized screen capture of desktop activity, and audio from the user's verbal protocol as shown in Figure 5. Students watched the video one time while the eye tracking equipment was turned on and identified critical incidents by pressing the space bar. After watching the video with eye tracking, students could then control playback of the video (play, pause, forward, reverse) while they logged what they perceived as usability problems on a paper-based critical incident form. During the rest of the semester, students were exposed to formal HCI concepts in the course. Near the end of the semester, students were assessed again using the same procedures and video clip used at the beginning of the semester.

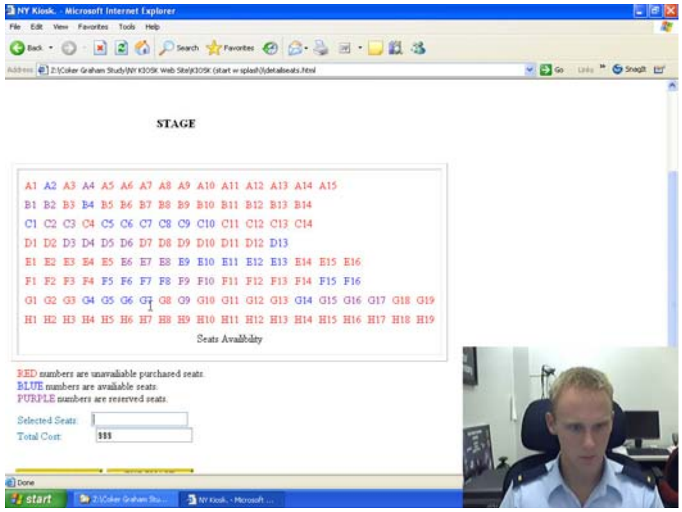

Figure 5: Sample clip of pre-recorded session of user interacting with web application

Results

We collected several measures during both the pre- and post-assessments in order to quantify any changes experienced by the students as they logged usability problems from the same video clip. These measures included attention focus (desktop vs. PIP), # of problems identified, word count, and use of HCI terms in describing usability problems.

Attention Focus

Attention focus involved the amount of time the students looked at the desktop activity vs. the PIP video of the user. Figure 6 shows the percent focus on desktop vs. PIP video in pre- and post-assessments. Pre- and post-assessments showed that students looked at the desktop activity approximately the same amount of time (80.74% vs. 81.96%). The same was true for their percent focus on the PIP video (12.09% vs. 12.02%). Because of their approximate similarity, the changes between pre- and post-assessments were not significant ($p > .10$).

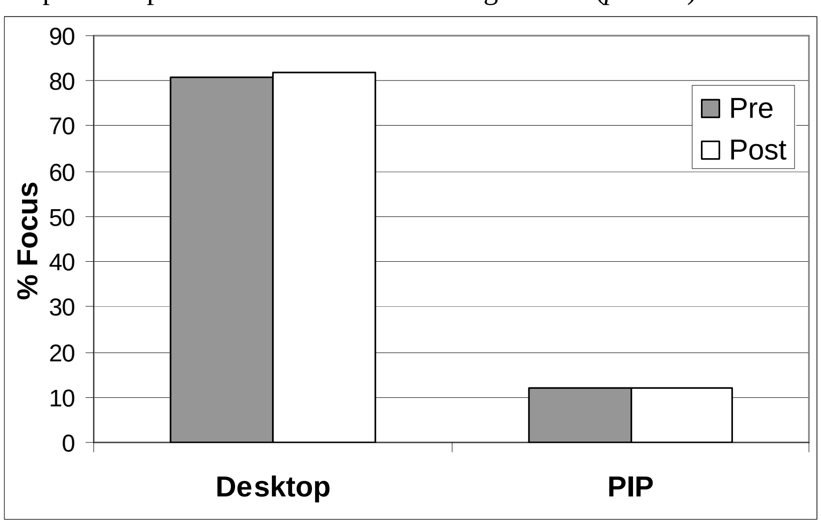

Figure 6: Percent focus on desktop activity vs. PIP video in pre- vs. post-assessments

Problems Identified

Students found an average of 13.38 problems in the pre-assessment and 17.15 problems in the post-assessment as shown in Figure 7. According to a paired-samples t-test, these results were significant [$t(12) = 4.003$, $p = .002$] indicating students did find more usability problems when they watched the video clip at the end of the semester. The standard deviation did not increase significantly from pre- to post-assessments (3.31 vs. 3.89).

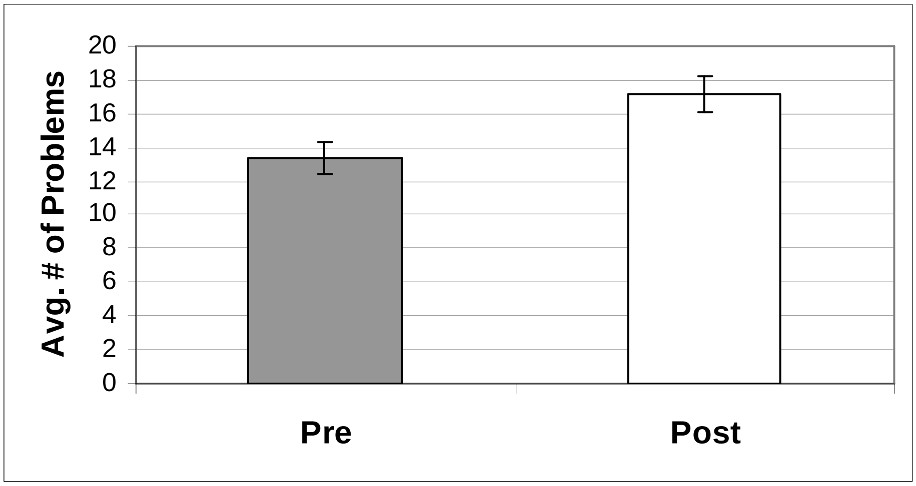

Figure 7: Average # of problems, pre- vs. post-assessment (+/- 1SE)

Word Count

We also looked at the number of words students used to log the usability problems they observed. To account for the different number of problems found in pre- vs. post-assessments, we used the average word count per problem (of each student) when describing usability problems. According to a paired-samples t-test, students used approximately the same number of words on average in pre- (10.90 words) vs. post-assessment (10.81 words), $t(12) = 0.07$, $p > .10$, to describe usability problems.

Use of HCI Terms

Taking a list of key words in HCI from Norman (2002) and Preece et al. (2002), we examined how frequently students used these words in their description of usability problems. These terms included feedback, visibility, affordance, conceptual model, and mapping to name a few. On average for each problem description, students used 4.61 HCI terms in the pre-assessment and 7.69 HCI terms in the post-assessment. A paired-samples t-test showed this difference to be moderately significant, $t(12) = 1.97$, $p < .10$.

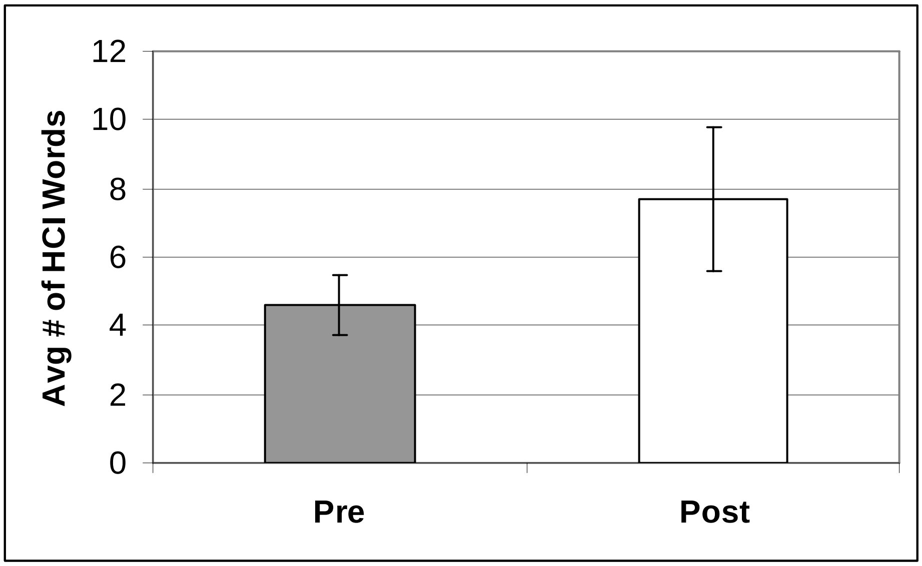

Figure 8: Average # of HCI words, pre- vs. post-assessment (+/- 1SE)

Discussion

Our objective in this study was to examine if changes in undergraduate student's usability technique could be measured over the course of a semester. Results showed that as they gained experience, students found more usability problems in pre- vs. post-assessment. Their attention focus remained the same, spending about the same amount of time looking at the PIP video of the user in pre- vs. post-assessments. Average word count for usability problem descriptions remained relatively the same over time. Most interesting, students appeared to use more HCI technical terms later in the semester.

Results from this study show that it is possible to quantify the usability evaluation experience differences of undergraduate students with some measures (e.g., number of problems identified and number of HCI technical terms). Future studies will include a larger sample of students and other objective measures that may show technique differences (e.g., eye scan patterns, number of fixations, and finding the most important problems).

Conclusion

The two studies conducted in this IITA project demonstrated the ability to use the Air Force Academy HCI laboratory in a way that allows for quantification of the practice of usability evaluation. Quantifying the effectiveness of usability evaluation methods provides guidelines for practitioners who use these specific methods to evaluate the results from usability studies. We have also been able to document how new students of HCI theory and methods are able to show specific behavioural changes as they gain experience in the discipline. Future work in the HCI laboratory will examine how to

evaluate and quantify usability problems in multi-user interface applications such as command and control environments.

References

Andre, T. S., Graham, H. D., Coker, J. L., & Schurig, M. A. (2006). Eye tracking of evaluators viewing usability videos: Opacity and experience differences. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Andre, T. S., Hartson, H. R., Belz, S. M., & McCreary, F. A. (2001). The user action framework: A reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies, 54(1)*, 107-136.

Argyle, M. (1972). Non-verbal communication in human social interaction. In R. A. Hinde (Ed.), *Non-verbal communication*. Cambridge: Cambridge University Press.

Argyle, M., & Dean, J. (1965). Eye-contact, distance and affilation. *Sociometry, 28*, 289-304.

Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in a dyad. *Human Relations, 21*, 3-17.

Gray, W. D., & Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction, 13*(3), 203-261.

Goldberg, J. H. and Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In J. Hyönä, R. Radach, & H. Deubel. (Eds.), *The mind's eye: Cognitive and applied aspects of eye movements research* (pp. 493-516). Amsterdam: Elsevier Science.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics, 24*, 631-645.

Hartson, H. R., Andre, T. S., & Williges, R. C. (2003). Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction, 15*(1), 145-181.

Kendon, A. (1967). Some function of gaze-direction in social interaction. *Acta Psychologica*, 26, 22-63.

Lin, Y. and Zhang, W. J. (2003). Evaluating interface usability based on eye movement and hand movement behavioral parameters. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 653-657). Santa Monica, CA: Human Factors and Ergonomics Society.

Long, K. M., Styles, L. J., Andre, T. S., & Malcom, W.C. (2005). Usefulness of nonverbal cues from participants in usability testing sessions. In G. Salvendy (Ed.), *Proceedings of the Human-Computer-Interaction International Conference* (CD ROM Vol. 4). St. Louis, MO: Mira Digital Publishing.

Norman, D. A. (2002). *The design of everyday things*. New York: Basic Books.

Patterson, M. L. (1983). *Nonverbal behavior: A functional perspective*. New York City, NY: Springer-Verlag.

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Beyond human-computer interaction*. New York: John Wiley & Sons.

Sears, A. (1997). Heuristic walkthroughs: Finding the problems without the noise. *International Journal of Human-Computer Interaction, 9*(3), 213-234.

About the Authors

Lt Col Terence Andre is the Deputy Department Head for Research in the Department of Behavioral Sciences and Leadership at the United States Air Force Academy. He directs the department's $1.2M laboratory infrastructure and research program for faculty and cadets. Lt Col Andre is an assistant professor in the department, teaching courses in human factors, system design, and human-computer interaction and oversees the human-computer interaction laboratory. He also directs cooperative agreements with outside research agencies. Lt Col Andre received his commission in 1987 after graduating from the United States Air Force Academy. He entered the Air Force in 1987 at Williams AFB, where he began his initial work as a human factors scientist. He had tours at Vandenberg AFB and Kirtland AFB as a program manager and human factors evaluator. He was recently the branch chief for the Warfighting Training Research Division at the Air Force Research Laboratory in Mesa, AZ where he led the division's $10M research program in Distributed Mission Training. Lt Col Andre received his Masters in Industrial Engineering from Cal Poly and his PhD in Industrial and Systems Engineering from Virginia Tech.

Margaret Schurig is a Human Factors Design Specialist for the Boeing Company. She is contracted through Air Force Research Laboratory (AFRL) in Mesa, Arizona to work at the United States Air Force Academy (USAFA) to assist in various research projects. Additionally, she develops websites, protocols, and tests new software for research purposes at AFRL. Ms. Schurig received her Bachelors of Science in Human Factors from Arizona State University in 2003.

About the Institute

The Institute for Information Technology Applications (IITA) was formed in 1998 to provide a means to research and investigate new applications of information technology. The institute encourages research in education and applications of the technology to Air Force problems that have policy, management, or military importance. Research grants enhance professional development of researchers b y providing opportunities to work on actual problems and to develop a professional network.

Sponsorship for the Institute is provided by the Assistant Secretary of the Air Force (Acquisition), the Air Force Office of Scientific Research, and the Dean of Faculty at the U.S. Air Force Academy. IITA Coordinates a multidisciplinary approach to research that incorporates a wide variety of skills with cost-effective methods to =achieve significant results. Proposals from the military and academic communities may be submitted at any time since awards are made on a rolling basis. Researchers have access to a highly flexible laboratory with broad bandwidth and diverse computing platforms.

To explore multifaceted topics, the Institute hosts single-theme conferences to encourage debate and discussion on issues facing the academic and military components of the nation. More narrowly focused workshops encourage policy discussion and potential solutions. IITA distributes conference proceedings and other publications nation-wide to those interested or affected by the subject matter.